

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO
FAKULTETA ZA MATEMATIKO IN FIZIKO

Gregor Mubi

Topološka analiza vremenskih slik

DIPLOMSKO DELO
UNIVERZITETNI ŠTUDIJSKI PROGRAM PRVE STOPNJE
RAČUNALNIŠTVO IN MATEMATIKA

MENTOR: prof. dr. Neža Mramor Kosta

Ljubljana 2015

Fakulteta za računalništvo in informatiko podpira javno dostopnost znanstvenih, strokovnih in razvojnih rezultatov. Zato priporoča objavo dela pod katero od licenc, ki omogočajo prosto razširjanje diplomskega dela in/ali možnost nadaljnje proste uporabe dela. Ena izmed možnosti je izdaja diplomskega dela pod katero od Creative Commons licenc <http://creativecommons.si>

Morebitno pripadajočo programsko kodo praviloma objavite pod, denimo, licenco *GNU General Public License*, različica 3. Podrobnosti licence so dostopne na spletni strani <http://www.gnu.org/licenses/>.

Besedilo je oblikovano z urejevalnikom besedil L^AT_EX.

Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Tematika naloge:

V diplomskem delu raziščite možnost uporabe topoloških metod, predvsem vztrajnosti povezanih komponent v podatkih, pri analizi radarskih vremenskih slik. Proučite možnost njihove uporabe pri razpoznavanju razmer za nastanek toče in drugih ekstremnih vremenskih pojavov in njihovo napovedovanje.

Pri predstavitvi radarskih vremenskih slik si pomagajte s strukturo kubičnega kompleksa. Radarsko sliko lahko predstavimo kot zaporedje naraščajočih kubičnih kompleksov, kjer posamezen kompleks zajema območje na sliki z vrednostjo funkcije, ki meri koncentracijo vodnih kapljic, nad določenim pragom. Na posamezni sliki lahko z metodo vztrajnosti zasledujemo, kako se s padajočim pragom funkcijske vrednosti pojavljajo nove povezane komponente kompleksa in kako se med seboj povezujejo. V časovnem zaporedju slik pa lahko zasledujemo dinamiko nastajanja in povezovanja komponent skozi čas. Poiščite ustrezne metode za vizualizacijo tako pridobljenih informacij in določite parametre, ki bi bili lahko uporabni pri modeliranju in napovedovanju izjemnih vremenskih pojavov.

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Gregor Mubi sem avtor diplomskega dela z naslovom:

Topološka analiza vremenskih slik

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom prof. dr. Neže Mramor Kosta,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela,
- soglašam z javno objavo elektronske oblike diplomskega dela na svetovnem spletu preko univerzitetnega spletnega arhiva.

V Ljubljani, dne 15. junija 2015

Podpis avtorja:

Za izdelavo diplomske naloge bi se želel zahvaliti Agenciji republike Slovenije za okolje, še posebej g. Antonu Zgoncu, za pomoč pri razumevanju podatkov, ki jih posname radar. Prav tako se zahvaljujem za poslane podatke, ki so mi omogočili testiranje algoritmov.

Zahvalil bi se tudi rad prof. dr. Neži Mramor Kosta za pomoč pri izvedbi diplomske naloge, predvsem pri razlagi teorije ter posredovanju virov in literature, kar mi je bilo v veliko pomoč.

Na koncu bi se še zahvalil asistentu as. dr. Juretu Žabkarju, ki je bil pobudnik za temo diplomskega dela in mi je podal veliko zanimivih idej in smernic.

Kazalo

Povzetek

Abstract

| | | |
|----------|------------------------------------------------------------|-----------|
| 1 | Uvod | 1 |
| 2 | Opis in obdelava vremenskih podatkov | 3 |
| 2.1 | Dekodiranje podatkov | 5 |
| 2.2 | Napake in šum podatkov | 6 |
| 3 | Teorija in razlaga osnovnih pojmov | 7 |
| 3.1 | Kubični kompleksi | 7 |
| 3.2 | Vztrajnost | 9 |
| 4 | Vztrajnost povezanih komponent na vremenskih slikah | 13 |
| 4.1 | Povezane komponente na slikah | 13 |
| 4.2 | Kdaj pride do toče? | 17 |
| 4.3 | Vztrajni diagrami visokih vrednosti | 17 |
| 4.4 | Dvodimenzionalna vztrajnost povezanih komponent | 21 |
| 5 | Sklep | 25 |
| | Literatura | 27 |

Povzetek

V nalogi je predstavljena analiza vremenskih podatkov s pomočjo računske topologije. Cilj naloge je analizirati vremenske slike in razbrati, ali lahko izpostavimo nekatere parametre, ki bi izboljšali klasifikacijo vlažnih gmot. V ta namen smo skozi zaporedje slik zasledovali, kako se rojevajo in združujejo vlažne gmote nad področjem na sliki. Uporabili smo topolosko metodo, ki se imenuje vztrajnost. Za analizo podatkov smo pridobili vremenske radarske slike. Za računanje in vizualizacijo podatkov smo napisali približno 20 skript v delovnem okolju MATLAB. Slike smo predstavili kot kubične komplekse, kjer so sivinske vrednosti podane v ogliščih kompleksa. Eden od rezultatov dela je tudi identifikacija parametrov, ki temelji na topoloških metodah in bi bil lahko eden od kazalnikov za napovedovanje toče in drugih izjemnih vremenskih pojavov.

Ključne besede: računska topologija, vztrajna homologija, kubični kompleksi, vztrajni diagrami, vremenski podatki, napovedovanje vremena.

Abstract

In this thesis computational topology is used to analyse weather data. The goal is to analyse weather images and discern, if it is possible to expose certain parameters, which could improve humid mass classification. For this reason we were observing how those masses appear and merge on sequences of images. We used topological method named persistence. For data analysis we obtained weather radar images. We also wrote around 20 scripts in computing environment MATLAB, which served for data computation and visualization. Images are presented as cubical complexes, where grayscale values are in vertexes of the complexes. One of the results is also an identification of parameter, which is based on topological methods and could be used in hail forecast and other exceptional weather forecasts.

Keywords: computational topology, persistent homology, cubical complexes, persistent diagrams, weather data, weather prediction.

Poglavje 1

Uvod

Od nekdaj smo ljudje želeli vedeti več o vremenu in čim bolj natančno napovedovati prihajajoče vremenske razmere. V meteorologiji se za napovedovanje vremena uporabljajo kompleksne metode in algoritmi, ki so prostorsko in časovno zelo zahtevni.

V tem diplomskem delu smo za analizo vremenskih podatkov uporabili topološko analizo podatkov - novejše področje, ki se je v zadnjih letih razvilo v okviru računske topologije. Topološka analiza podatkov v zadnjem času postaja vse bolj popularna, saj daje poudarek vizualnim lastnostim podatkov (na primer številu povezanih delov in številu lukenj v podatkih).

Za izvedbo željene analize smo najprej pridobili vremenske podatke. Podatke nam je posredovala Agencija Republike Slovenije za okolje (ARSO). Podatki zajemajo časovna zaporedja radarskih slik nad Slovenijo in njeno okolico. Te smo uporabili za nadaljno analizo.

Podatke smo analizirali z različnimi topološkimi metodami, ki so bile sprva preprostejše, kasneje pa vse bolj kompleksne in informativne. Kot najbolj uporabni so se pokazali vztrajni diagrami. Za konstrukcijo vztrajnih diagramov smo uporabili različne filtre, ki omogočajo vpogled v podatke pri različnih resolucijah. V prvem delu so predstavljeni bolj sploši rezultati, na primer število povezanih komponent na sliki in njihova vztrajnost. V drugem delu pa smo se osredotočili na visoke vrednosti funkcije in primere ekstremnih

odbojnosti, ki bi lahko predstavljale tudi točo.

Cilj naloge je bilo ugotoviti, ali lahko s pomočjo topoloških metod, predvsem vztrajne homologije, pridobimo neke nove informacije in parametre za napovedovanje vremena ter klasifikatorje za boljše prepoznavanje vremenskih pojavov.

Poglavje 2

Opis in obdelava vremenskih podatkov

Agencija RS za okolje (ARSO) nam je posredovala podatke o radarskih slikah nad območjem Slovenije in njeno okolico, ki jih pridobivjo z dvema radarjema. Podatki, ki smo jih analizirali v tem delu, opisujejo odbojnost radarskega signala nad vsako točko na zajetem območju. Ta predstavlja količino vodnih kapljic v zračnem stolpcu nad izbrano točko. Odbojnost žarka je merjena v decibelih glede na refleksivnost Z (decibels relative to Z - DBZ) in je zapisana v tekstovni datoteki v obliki 7-bitnih ASCII znakov. Radar naredi vsakih 10 minut eno sliko, to je 144 slik vsak dan.

Vrednosti se pričnejo z 0 dBZ in rastejo do okoli največ 75 dBZ. Vrednosti pod 15 dBZ so komaj opazne in jih zanemarimo zaradi premajhnega vpljiva na opazovano stanje. Najbolj zanimive so vrednosti nad 55 dBZ, ki predstavljajo nevarnejša stanja, kot je toča. Te vrednosti je smiselno spremljati in ugotavljati, kakšni so pogoji za njihov nastanek.

Primerjava med vrednostmi dBZ, milimetri padavin na uro (mm/h) in intenziteto padavin je podana v naslednji tabeli. Tabela zajema vrednosti med 5 in 65 dBZ, saj v ta interval pade več kot 99% odčitanih podatkov.

| dBZ | R (mm/uro) | Intenziteta |
|-----|------------|-----------------------------------|
| 5 | 0.07 | komaj opazno |
| 10 | 0.15 | rahla megla |
| 15 | 0.3 | megla |
| 20 | 0.6 | zelo rahle padavine |
| 25 | 1.3 | rahle padavine |
| 30 | 2.7 | rahlo do zmerne padavine |
| 35 | 5.6 | zmerne padavine |
| 40 | 11.53 | zmerne padavine |
| 45 | 23.7 | zmerne do močnejše padavine |
| 50 | 48.6 | močnejše padavine |
| 55 | 100 | zelo močne padavine / manjša toča |
| 60 | 205 | zmerna toča |
| 65 | 421 | velika toča |

Tabela 2.1: *Razmerje med vrednostmi dBZ in padavinami*

Iz Tabele 2.1 je razvidno, da je razmerje med vrednostjo dBZ in milimetrih padavin na uro približno kvadratično in ne linearno. To pomeni, da so vrednosti dBZ bolj občutljive na manjše padavine in te bolj natančno med seboj ločuje.

2.1 Dekodiranje podatkov

Podatki Agencije RS za okolje o radarskih slikah so na voljo v tekstovni datotetki. Datoteka vsebuje na začetku 25 parametrov, ki jih vpiše radar. Med bolj uporabnimi so na primer:

- dimenzija slike (v konkretnem primeru 401x301),
- datum in čas slike,
- oznaka radarja, ki je slikal (prvi, drugi ali oba),
- velikost celie,
- število nivojev odbojnosti (v konkretnem primeru 16),
- začetna vrednost računanja odbojnosti,
- širina (razlika med vrednostmi),
- vrednost, kjer radar ni imel dostopa do podatkov,
- središče radarja,
- enota, v kateri so predstavljeni podatki (v konkretnem primeru dBZ).

Temu sledijo podatki, ki so predstavljeni z znaki ASCII tabele. Vsak znak predstavlja vrednost odbojnosti na sliki (npr. @ = 12dBZ). Če želimo pretvoriti znake s tekstovnega dokumenta v vrednosti odbojnosti, moramo upoštevati zgoraj omenjene parametre. Te lahko združimo v spodnji račun, ki pove vrednost odbojnosti v enem slikovnem elementu (pikslu):

$$\text{začetna vrednost} + (\text{vhodni znak} - \text{zamik}) * \text{širina} = \text{vrednost odbojnosti}$$

Sledi komentar v vrstici, ki se začne z znakom `#`. Tu so pogosto podatki, ki povedo, kako in kdaj je datoteka nastala, kateri vhodni podatki so bili uporabljeni in kdo je ustvaril datoteko.

Skripta **ProcessDataTable.m** izlušči vrednosti funkcije odbojnosti in jih zapiše v tabelo v formatu, ki je primeren za obdelavo z MATLABom. Na vhod sprejme ime datoteke in vrne tabelo dimezij 401 x 301, kar predstavlja širino in višino slike, ki jo posname radar v slikovnih elementih. Uporabili smo tudi skripto **TimeToFileName.m**, ki skrbi za pretvorbo številke od 1 do 144 v ime datoteke (na primer: *si0-20140625-1810-zm.srd*).

2.2 Napake in šum podatkov

V pridobljenih podatkih je kar nekaj šuma in motenj. Najbolj izraziti primeri so interferenca, sence in ločljivost. Do interference pride zaradi dveh radarjev, kjer prihaja do sferičnih motenj, ki se na sliki vidijo kot krogi povečanih odbojnih vrednosti. Ločljivost zajetih podatkov pada z oddaljevanjem od radarjev. Visoke gore preprečujejo radarjem zajemati podatke za njimi. Ta pojav se na sliki vidi kot senca.

Teh napak smo se pri analizi podatkov zavedali, vendar smo se odločili, da ne bomo spreminjali vhodnih podatkov, saj se topološki pristop osredotoča na obliko podatkov in menimo, da te napake ne vplivajo na končen rezultat. To smo na sliki 4.7 tudi pokazali. Na sliki je razvidna napaka v podatkih, vendar ta napaka ne vpliva na končen rezultat.

Poglavje 3

Teorija in razlaga osnovnih pojmov

V diplomskem delu želimo analizirati vremenske podatke na nek nov način. Odločili smo se, da poizkusimo z novejšim pristopom, ki se vse bolj pogosto uporablja v analizi podatkov, to je topološki pristop. Za ta pristop je značilno, da je v ospredju oblika in značilnosti oblike neke množice točk. Na začetku smo si ogledali več splošnih topoloških idej, vendar smo se na koncu osredotočili le na dvodimenzionalne kubične komplekse in vztrajnost povezanih komponent, saj so nam dajali boljše rezultate. Pri razlagi pojmov in definicij smo si pomagali s spletnim učbenikom Dr. Petra Saveliev [2] in knjigo Computational topology Herberta Edelsbrunner [5]. Za boljšo predstavitev pojmov smo uporabili tudi nekaj slik iz virov, navedenih v literaturi.

3.1 Kubični kompleksi

Vremenske slike so v diplomskem delu predstavljene s kubičnimi kompleksi. Omejili smo se na dvodimenzionalne komplekse, saj so vhodni podatki prav tako dvodimenzionalne slike.

Definicija 1. Kocka ali celica dimenzije n je kartezični produkt n intervalov:

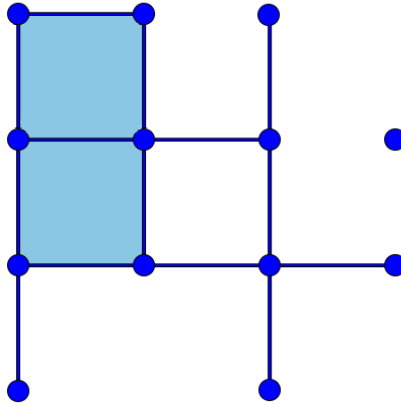
$$\sigma^n = [a_1, b_1] \times \cdots \times [a_n, b_n]. \quad (3.1)$$

Kocke nižje dimenzije σ^k , ki so v σ^n , predstavljajo lica kocke σ^n .

Definicija 2. Kubični kompleks K dimenzije N je sestavljen iz kock dimenzij $0, 1, \dots, N$, ki so pravilno sestavljene, tako da velja:

1. za vsako kocko v kubičnem kompleksu K , so vsa njena lica tudi vsebovana v K ;
2. presek dveh kock iz kubičnega kompleksa K je skupno lice obeh.

Za vsako kočko σ v kubičnem kompleksu K , $\sigma \in K$ imamo v kompleksu tudi vsa njena lica. Presek poljubnih dveh kock je skupno lice.



Slika 3.1: Dvodimenzionalni kubični kompleks je sestavljen iz oglišč (kock dimenzije 0), robov (kock dimenzije 1) in kvadratov (kock dimenzije 2).

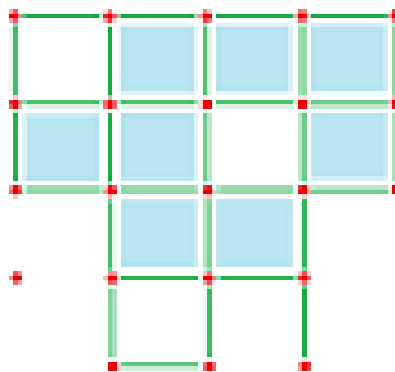
Kubični kompleks predstavimo s seznamom kock, ki ga sestavljajo, in za vsako kocko so navedeni podatki o tem, katere kocke nižje dimenzije tvorijo

njen rob. Ti podatki so običajno shranjeni v robnih oz. incidenčnih matrikah, kjer stolpci ustrezajo kockam dimenzije n , vrstice pa kockam dimenzije $n - 1$, $n = 1, \dots, N$. Na mestu (i, j) je 0, če kocka σ_i^{n-1} ni v robu kocke σ_j^n in 1 če je.

3.2 Vztrajnost

Kubični kompleks lahko razpade na več nepovezanih kosov, ki jim pravimo povezane komponente.

Definicija 3. Kubični kompleks K je *povezan*, če lahko poljubni oglišči $a, b \in K$ najdemo zaporedje robov, ki ju povezuje.



Slika 3.2: Primer kubičnega kompleksa, ki ni povezan, ker spodnjega levega oglišča ne moremo z zaporedjem robov povezati z ostalimi oglišči.

Poljuben kubični kompleks K razpade na unijo manjših, povezanih kubičnih kompleksov, ki jim pravimo *povezane komponente* kompleksa K . Na primer, kubični kompleks na sliki razpade na dve povezani komponenti - ena vsebuje samo spodnje levo oglišče, druga pa vse ostalo.

V našem primeru so podatki dvodimenzionalne radarske slike, kjer so oglišča posamezni slikovni elementi (piksli). Dve oglišči sta povezani, če

sta si slikovna elementa na sliki sosedna v vodoravni ali v navpični smeri. V tem primeru ju v kubičnem kompleksu povežemo z robom. Kvadrati v kubičnem kompleksu ustrezajo sosednim slikovnim elementom, ki na sliki tvorijo kvadrat. Na dvodimenzionalni sliki ima tako lahko vsako oglišče od nič do štiri povezave s sosednimi oglišči in je lahko oglišče največ štirih kvadratov.

Na radarskih slikah bomo zasledovali, kako se tvorijo vlažne zračne gmote, ki predstavljajo oblake, in se skozi čas združujejo in razdružujejo. V ta namen kubični kompleks, ki ustreza pozamezni radarski sliki, zgradimo postopno, tako da upoštevamo čedalje več slikovnih elementov glede na njihovo stopnjo odbojnosti. Dobimo zaporedje čedalje večjih vgnezenih kubičnih kompleksov, ki mu pravimo filtracija.

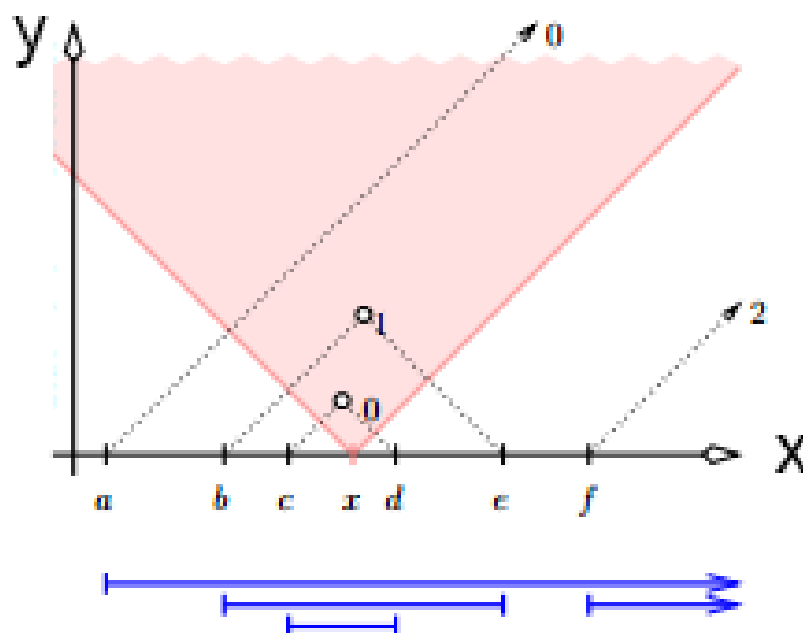
Definicija 4. Filtracija kubičnega kompleksa je zaporedje, kjer začnemo s prazno množico in postopoma dodajamo elemente kubičnega kompleksa K .

$$\emptyset \subset K_0 \subset K_1 \subset K_2 \subset \dots \subset K_n = K, n \in \mathbb{Z} \quad (3.2)$$

Postopek, ki zasleduje nastajanje in združevanje povezanih komponent skozi filtracijo, imenujemo *vztrajnost*.

Ko kubični kompleks, ki predstavlja radarsko sliko, skozi filtracijo narašča, se pojavljajo nove povezane komponente, ki se med sabo združujejo v večje povezane komponente. Če se ob stopnji j dve povezani komponenti združita, se ohrani starejša, mlajša pa izgine (umre). Za vsako povezano komponento beležimo stopnjo njenega rojstva i in stopnjo njene smrti j . Za povezane komponente, ki vztrajajo do konca in na koncu predstavljajo povezane komponente kubičnega kompleksa K , je stopnja smrti ∞ .

Vsaki povezani komponenti, ki nastane med filtracijo, pripada par (i, j) , kjer je i stopnja rojstva med 0 in n , in j stopnja smrti med 0 in ∞ . Pare (i, j) lahko vizualno predstavimo na dva načina: s pomočjo *črtne kode* (*barcode*) [6, 7] ali s pomočjo *vztrajnega diagrama* (*persistence diagram*) [8].



Slika 3.3: Vztrajni diagram in pripadajoče črtne kode.

Na sliki 3.3 je predstavljen zavrteni vztrajni diagrami, kjer so koordinate točke $(i+j, i-j)$. Na x osi je vsota stopenj, kjer se povezana komponenta pojavi in kjer izgine. Na y osi je razlika stopenj, kjer komponenta izgine in kjer se pojavi. Iz te oblike lahko dobimo drugi način, kako predstavljamo vztrajnost povezanih komponent, ki mu pravimo *črtne kode*. Na sliki so predstavljene kot modre črte pod grafom. Dobimo jih tako, da naredimo poševni projekciji točke na x os. Črtne kode so intervali z začetkom v času, kjer komponenta nastane, in se končajo ob času, kjer komponenta izgine. V primeru, da komponenta v opazovanem času ne izgine, interval na grafu označimo s puščico, ki nakazuje nadaljevanje komponente. Dolžina črtne kode nam tako pove, koliko časa je povezana komponenta obstajala.

Za izris vztrajnih diagramov in črtnih kod smo si tudi pomagali s knjižnico JavaPlex. Ta na vhod sprejme seznam točk, povezav in čas pri katerem se element pojavi ali izgine. Seznam točk ustvarimo z metodo

```
stream = api.Plex4.createExplicitSimplexStream(),
```

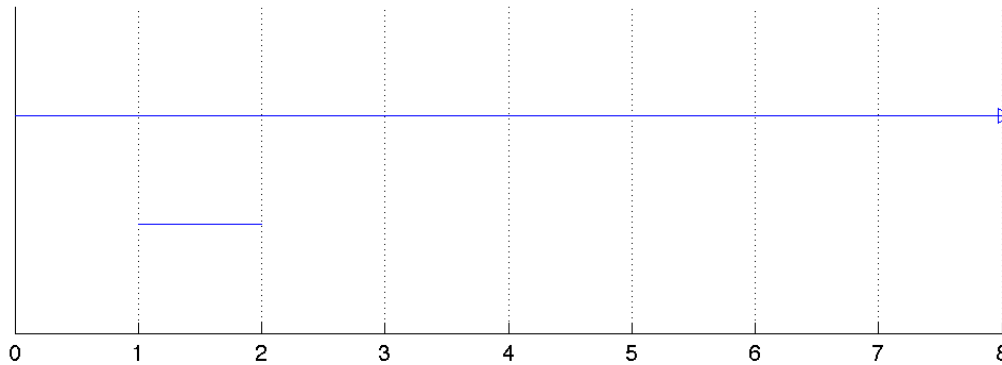
nato pa dodajamo točke z metodo

```
stream.addVertex(i, t),
```

pri čemer je i index točke, t pa čas. Nato lahko izpišemo intervala povezanih komponent z metodo

```
intervals = persistence.computeIntervals(stream).
```

Ta nam vrne seznam intervalov, ki jih lahko izrišemo z metodo `plot_barcodes(intervals, options)`.



Slika 3.4: Primer preprostih črtnih kod z uporabo metod iz knjižnice *javaPlex*.

Iz slike lahko razberemo, da v opazovanem času obstajata dve komponenti. Prva nastane v času $t = 0$ in traja v neskončnost. Druga se pojavi v času $t = 1$ in izgine v času $t = 2$.

Poglavje 4

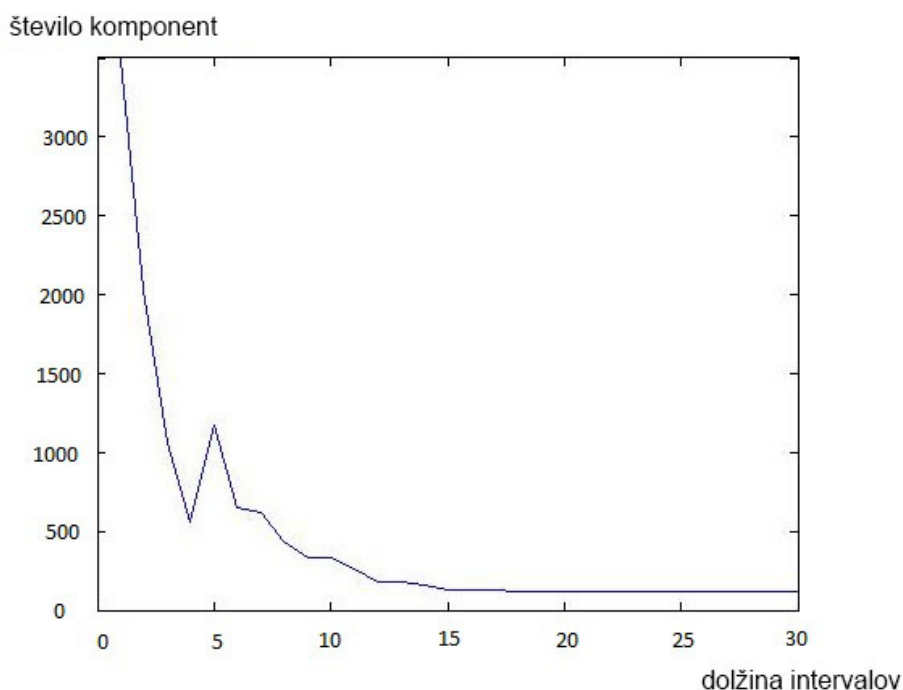
Vztrajnost povezanih komponent na vremenskih slikah

4.1 Povezane komponente na slikah

Na začetku smo se lotili izračunati število komponent na podlagi primera, ki je bil predstavljen na predavanjih Isabel K. Darcy [1]. Število povezanih komponent smo najprej izračunali tako, da smo vse točke na sliki razdelili v prekrivajoče se intervale glede na vrednost funkcije odbojnosti. Znotraj vsakega intervala smo med seboj povezali vsaki dve točki, ki sta si sosedni na radarski sliki. Tako smo dobili neko število povezanih komponent. Postopek smo ponovili večkrat, pri tem pa vedno povečali širino intervalov. Tako se je število povezanih komponent spreminjalo, saj so se spreminjale povezave med točkami.

Za ta namen smo spisali skripto **GetIntervalPersistence.m**, ki na vhod sprejme radarsko sliko, vrne pa graf, ki prikazuje razmerje med številom komponent in širino intervalov. Primer grafa lahko vidimo na spodnji sliki. Z grafa je razvidno, da število komponent pričakovano pada z dolžino intervalov.

Postopek je bil ponovljen na več zaporednih radarskih slikah. Z njihovim primerjanjem smo ugotovili, da so razlike zelo majhne, skoraj zanemarljive. Oblika grafa je na vseh slikah podobna. Vsaka slika ima nad tri tisoč komponent pri manjših dolžinah intervalov in med sto in dvesto komponent pri večjih dolžinah intervalov.

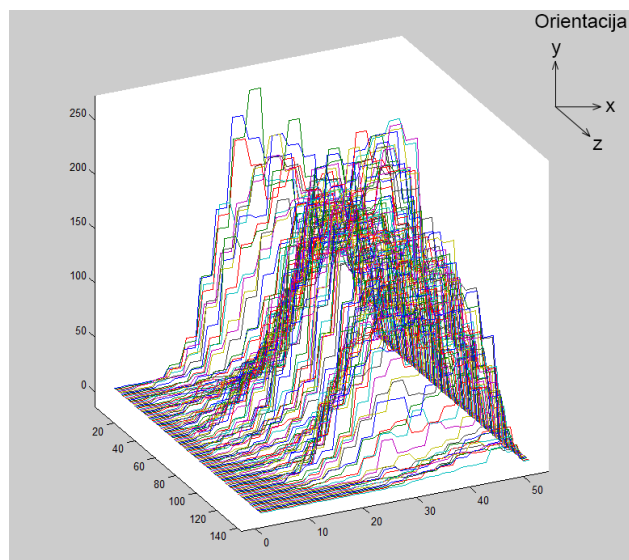


Slika 4.1: Graf prikazuje razmerje med številom komponent in dolžino intervalov.

Ugotovili smo, da izrisani grafi niso najbolj uporabni. V prejšnjem primeru smo razdelili točke v več intervalov glede na njihovo vrednost. Odločili smo se, da bomo filtrirali podatke glede na primerno izbran prag. Vse točke, ki imajo vrednost pod pragom, bomo odstranili s slike, ostale pa med seboj povezali, če so si sosedne. Prag lahko spreminjamo. Tako zopet dobimo razmerje med številom komponent in vrednostjo praga. To razmerje lahko prikažemo na grafu.

Postopek smo ponovili na vseh radarskih slikah enega dneva. Vseh 144

pridobljenih grafov slik smo nato izrisali drug poleg drugega, tako da smo dobili tridimenzionalno sliko.



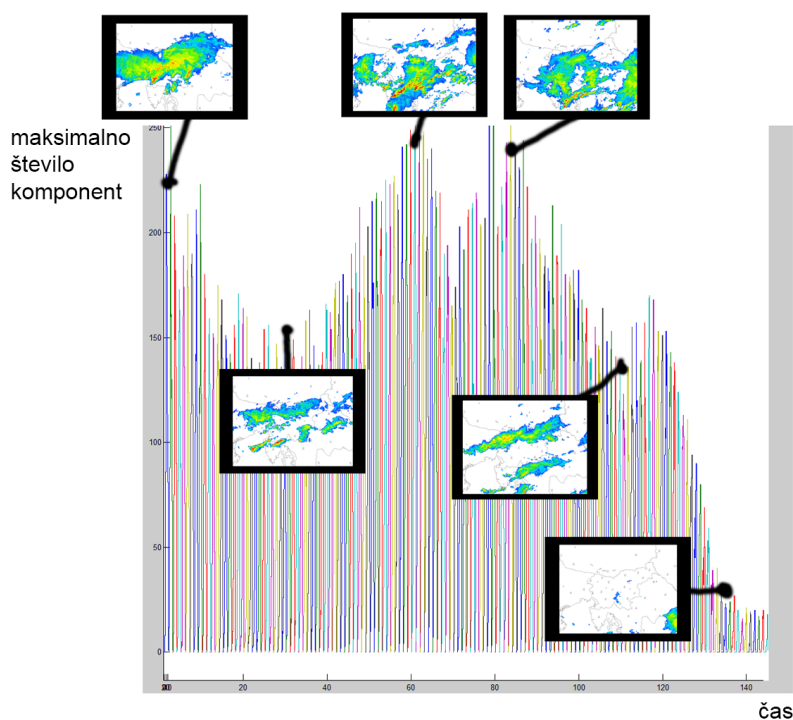
Slika 4.2: Graf vseh 144 razmerij med številom komponent in vrednostjo praga.

Graf prikazuje, da je bilo na izbrani dan prisotnih veliko vlažnih gmot. Pri proučevanju moramo biti pozorni na strme predele grafa, ki predstavljajo področja, kjer je prihajalo do večjih sprememb v relativno kratkem času.

Na sliki 4.2 lahko vidimo vrednost praga (števila od 0 do 50 na osi x), število komponent (števila od 0 do 250 na y osi) in zaporedno številko vhodne slike (števila od 0 do 144 na z osi). S te slike je možno razbrati več uporabnih podatkov. Najbolj zanimiva podatka sta maksimalno število komponent na osi y in razmerje med številom komponent in vrednostjo praga. Pri visoki vrednosti praga vidimo malo komponent, nato se število komponent poveča, za tem pa zmanjša, zaradi združevanja gmot.

Zanimivi podatki o vremenu, ki jih lahko vidimo na sliki, bi lahko bili prag pri katerem začne število komponent naraščati, območje z največ komponentami, prag, kjer število komponent stmo pade in spreminjanje teh parametrov preko dneva.

Zanimiv podatek je tudi število komponent na vsaki posamezni sliki. To lahko prikažemo, če naredimo projekcijo na ravnino osi y in z oziroma tridimenzionalno sliko pogledamo z leve ali desne strani. Prav tako so zanimive tiste slike, kjer v relativno kratkem času število komponent pade. Glede na maksimalne vrednosti odbojnosti bi to lahko pomenilo, da se je vreme nenadoma umirilo in se je vlažnost enakomerno razširila na celotno območje ali pa, da se se posamezne vlažne gmote hitro združile, kar je lahko znak za nastajajočo točo.



Slika 4.3: Maksimalna števila komponent za vsako sliko v dnevu.

Tako smo dobili graf števila komponent za vsako sliko v danem dnevu. Na zgornji sliki smo primerjali nekatere vrednosti z njihovimi radarskimi slikami. Ugotovili smo, da je maksimalno število komponent sorazmerno s pokritostjo radarske slike z vlažnimi gmotami.

Zanimivo bi bilo tudi upoštevati, pri kateri vrednosti funkcije je doseženo maksimalno število komponent. Iz tega bi lahko sklepali, kakšno je ozračje. Pri nizki vrednosti je bolj umirjeno ozračje, pri manjši vlažnosti se oblikujejo velike homogene gmote z relativno nizko vlažnostjo, pri višji vrednosti pa je ozračje bolj nehomogeno in turbulentno. Večje homogene gmote se oblikujejo šele pri visoki vlažnosti, ki bi morda lahko predstavljala težke nevihtne oblake.

4.2 Kdaj pride do toče?

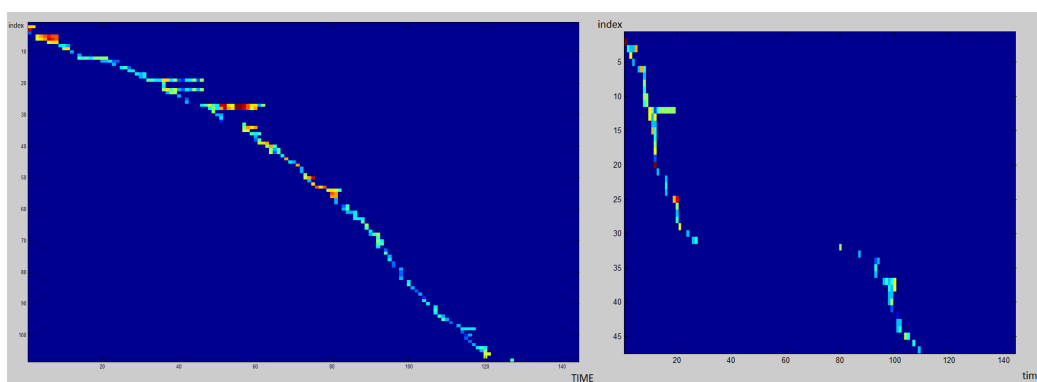
Želeli smo izvedeti kaj več o ekstremnih vremenskih pojavih. Na radarski sliki so ti predstavljeni z rdečo barvo, ki predstavlja višje odbojne vrednosti. Eden izmed ekstremnih vremenskih pojavov, ki so bolj nepredvidljivi, je zagotovo toča.

V teoriji je eden od pogojev za nastanek toče ta, da se relativno hitro zelo veliko vlage dvigne v zrak. Torej na slikah iščemo visoke vrednosti odbojnosti, ki so nastale v zelo kratkem času, se dvignile, omogočile nastanek toče in tudi hitro izginile. Takšne pogoje v radarskih slikah najdemo tako, da hkrati opazujemo vztrajnost povezanih komponent pri filtraciji slik glede na funkcijske vrednosti in vztrajnost povezanih komponent glede na časovno zaporedje slik. V črtnih kodah se ta pojav kaže kot dolge črte (predstavljajo veliko odbojnost) na vztrajnih diagramih s filtracijo glede na vrednost in kratke črte na vztrajnih diagramih s filtracijo glede na čas (predstavljajo hitro porast vlažnosti na pripadajočem območju). Iščemo torej dolge črte pri filtraciji slike po vrednosti funkcije in kratke črte v časovnih zaporedjih slik.

4.3 Vztrajni diagrami visokih vrednosti

Do toče in drugih ekstremnih pojavov pride le pri visokih vrednostih odbojnosti, zato smo iz slik odstranili dele, kjer je odbojnost nizka, saj nas ti ne zanimajo. V ta namen smo pripravili skripto **GetHailPersistence.m**. Tako

smo obdelali več zaporednih slik in jih med seboj primerjali. Z dveh zaporednih slik je razvidno, da so nad istim območjem radarske slike podobne odbojne vrednosti. Logično je, da je to ista vlažna gmota na obeh slikah, ki je rahlo spremenila obliko in se nekoliko premaknila. Tako lahko sledimo tem gmotam v času. Vsaka gmota ima nek trenutek, ko se pojavi in trenutek v katerem izgine ali združi z drugo starejšo gмотo. Tem gmotam smo sledili skozi čas in jih izrisali na naslednji sliki kot črtne kode.



Slika 4.4: Črtne kode visokih odbojnih vrednosti za dva različna dneva.

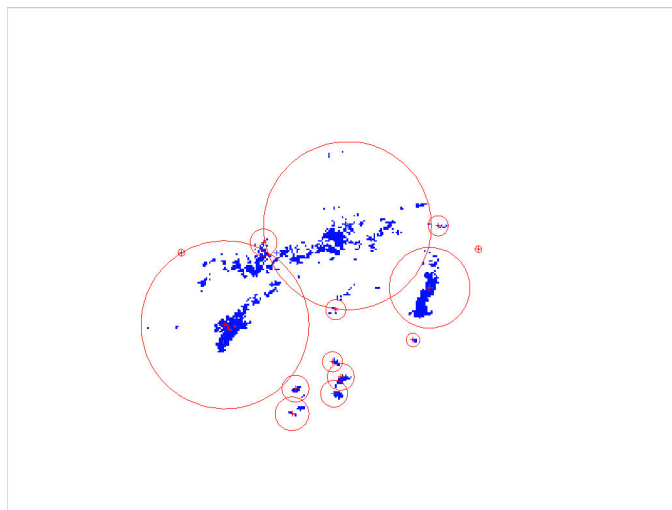
Zanimivi podatki na sliki so svetlejše črte na temno modri podlagi. Grafa pokazeta, koliko časa trajajo vlažne gmote z visoko vrednostjo odbojnosti. Daljša je horizontalna črta, dalj časa vztraja gmota. Barva predstavlja velikost gmote. Rumene in rdeče barve predstavljajo večje gmote, zelene in modre barve pa manjše. Vsaka novo opažena gmota ima svojo vrstico. Opazili smo veliko svetlo modrih in zelenih kratkih črt. Te predstavljajo majhne gmote, ki so prisotne kratek čas. Opazili smo tudi nekaj rumenih in rdečih črt, ki predstavljajo večje gmote.

Z grafa na levi je razvidno, da gre za konstantno prisotnost vlažnih gmot z visoko odbojno vrednostjo čez dan. Drugi graf ima nekaj gmot, ki trajajo kratek čas, nato izginejo in se proti koncu grafa spet pojavijo.

Nevarnost toče po hipotezi predstavljajo vlažne gmote, ki trajajo malo časa. Z zgornje slike je razvidno, da je takih črt veliko. Poleg informacije, ki jo vsebuje takšen graf, potrebujemo še informacijo, kako hitro se je vlažna

gmota pojavila in kako hitro je njena odbojna vrednost narasla.

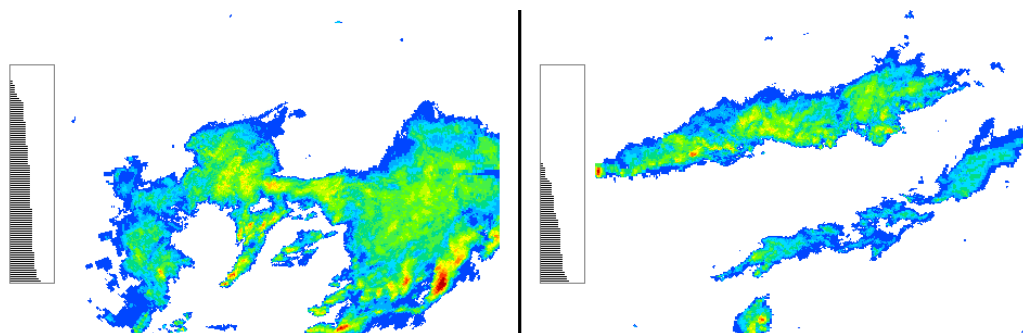
Do sedaj smo sledili vlažnim gmotam z visoko odbojno vrednostjo. Dobro pa bi bilo vedeti tudi, kako hitro so te gmote nastale. Zato smo radarsko sliko filtrirali s pragom od zgornjih vrednosti navzdol. Tako smo dobili zaporedje slik, na katerih je vse več točk, dokler ne zajememo celo prvotno sliko. Točke, ki so si sosedne, povežemo v eno komponento ter sledimo s slike v sliko povezanim komponentam, kako rastejo in se združujejo.



Slika 4.5: Primer označevanja potrebnega za izris vztrajnih diagramov.

Na sliki je prikazan primer slike pri eni izmed filtracij. Modra barva predstavlja vrednosti nad pragom in rdeči krogi približno opisujejo povezane komponente. Velikost kroga ustreza premeru komponente (največji razdalji med dvema točkama) in nakazuje velikost posameznih komponent (služi zgolj za vizualizacijo). Programska koda je v skriptah **GetHailPersistence.m** in **AddRadarPicToValuePersistence.m**. Skripti skrbita za označevanje in povezovanje podatkov med slikami skozi čas.

Zaporedju slik lahko priredimo črtne kode. Te nam povedo število komponent na dani sliki - kako hitro komponente nastanejo in koliko časa trajajo. V ta namen smo pripravili video, kjer smo izrisali radarske slike in črtne kode, ki smo jih pridobili s slike. Spodaj lahko vidimo dve sliki, ki sta del videa.



Slika 4.6: Radarski slike in pripadajoče črtne kode

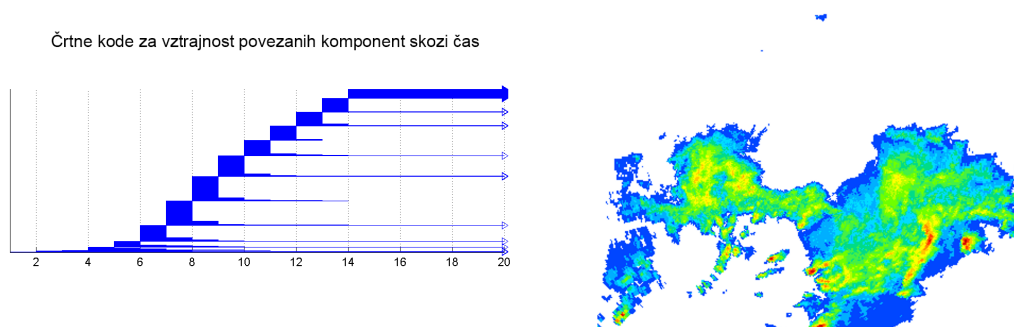
Črtne kode na levi predstavljajo pojavljane novih vlažnih gmot ob nižanju praga, pri čemer nismo upoštevali združevanja gmot. Iz oblike stolpca lahko razberemo nekatere lastnosti slike (na primer, če je stolpec na vrhu zabljen, potem je vlažnost ozračja bolj homogena).

Slike kažejo, da je število komponent sorazmerno s pokritostjo radarske slike z vlažnimi gmotami. Prav tako smo zbrali podatke za vse povezane komponente. Na teh podatkih bi sedaj lahko nadaljevali naše delo, vendar se pri tem algoritmu točke znotraj povezanih komponent povezujejo med seboj le znotraj radija, ki je omejen. Odločili smo se, da izboljšamo ta del algoritma.

V drugem delu smo za računanje in predstavitev uporabljali knjižnico javaPlex, ki se jo lahko naloži v programsko okolje Matlab. Knjižnica pospeši in poenostavi metode in funkcije, ki smo jih do sedaj znova in znova pisali in izpopolnjevali sami. Za uporabo metod in funkcij knjižnice smo se zgledovali z vira [4].

4.4 Dvodimenzionalna vztrajnost povezanih komponent

Zadnja vizualizacija vremenskih podatkov, ki smo jo razvili v diplomskem delu, predstavlja vztrajnost črtnih kod posameznih komponent. Za vsako radarsko sliko posebej smo imeli zbrane že od prej. Vsaka črtna koda na posamezni radarski sliki predstavlja eno povezano komponento. Nato smo vse radarske slike enega dneva postavili v zaporedje. Tokrat smo med seboj povezali črtne kode na različnih slikah, ki pripadajo isti povezani komponenti. Tako smo dobili črtne kode za povezane komponente v času. Te predstavljajo, koliko časa povezane komponente trajajo.



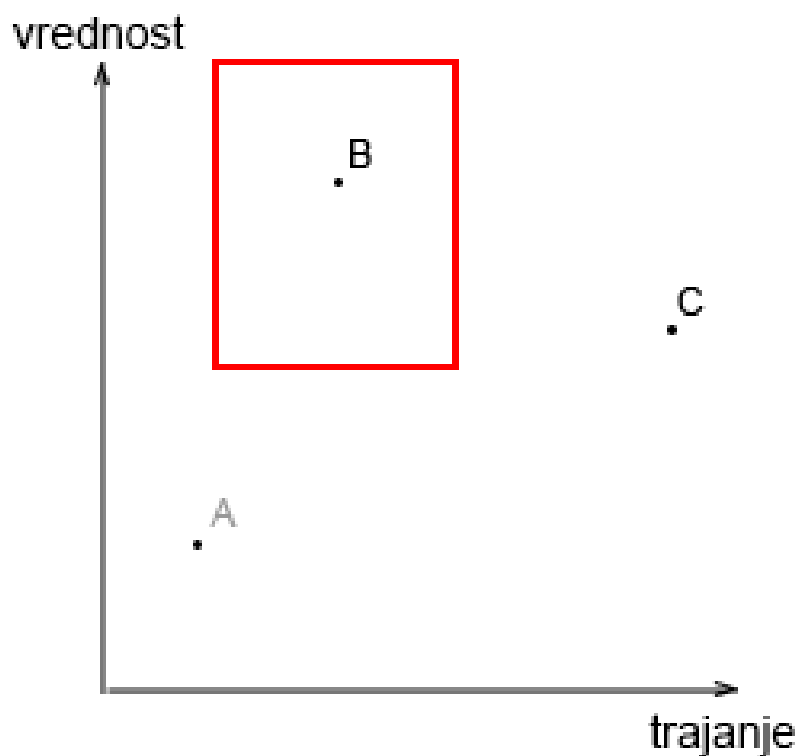
Slika 4.7: Črtne kode in radarska slika konkretnega primera.

Opomba: Stopničasta zgradba je posledica zapisa radarskih podatkov. Predstavljena slika je na nekaterih delih precej zrnata, kar vpliva na veliko število nepovezanih komponent.

S slike 4.7 je razvidno, da se ob vsakem spustu praga pojavi veliko novih povezanih komponent, ki se ob naslednjem spustu hitro združijo. Razlog za to je zrnatost slike. Z radarske slike na desni vidimo, da je več predelov, kjer se prepleta več barv (na primer zelena in rumena barva). Rumena predstavlja višje odbojne vrednosti kot zelena. V nekem koraku smo spustili prag do te vrednosti, da so rumene točke že del končne slike, zelene točke pa še ne. Tako nastane veliko rumenih točk, ki se med seboj ne morejo povezati, saj

smo zelene predele med njimi s filtrom odstranili. V naslednjem koraku se vrednost praga spusti in zapolni prazne prostore med rumenimi točkami in te se med seboj povežejo v manjše število komponent.

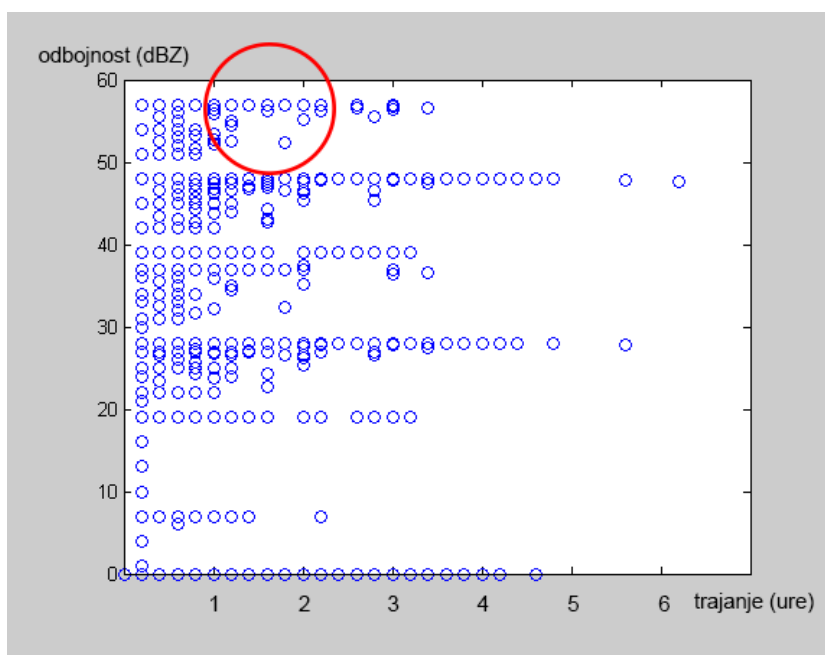
Vsako povezano komponento smo tako predstavili z dvema naboroma črtnih kod. S prvega nabora lahko razberemo odbojne vrednosti povezanih komponent, z drugega pa njihovo trajanje. Vsako povezano komponento smo predstavili kot točko $T(x,y)$ na dvodimenzionalnem grafu, kjer vrednost x predstavlja trajanje komponente, vrednost y pa njeno odbojno vrednost.



Slika 4.8: Način predstavitve rezultatov.

Na zgorji sliki smo označili tri zanimive vrste tako dobljenih vztrajnih točk:

- okolica točke A: sem spadajo povezane komponente, ki ne predstavljajo padavin, ampak oblake in meglice. Pojavijo se za krajši čas, na primer do nekaj ur;
- okolica točke B: v tej okolici najdemo komponente z visokimi odbojnimi vrednostmi, kar predstavlja močnejša deževja za krajši čas. Po hipotezi, bi primeri toče spadali v to okolico;
- okolica točke C: te komponente imajo prav tako višje odbojne vrednosti, kot okolica točke B, vendar trajajo veliko dalj časa. Dobri primeri za to okolico so daljša deževja. Z njimi bi lahko tudi povezali nekatere poplave.



Slika 4.9: Razporeditev povezanih komponent glede na trajanje in odbojnost.

Območje označeno z rdečim krogom po hipotezi pripada komponentam, ki so potencialni kandidati za pojavitev toče.

Sliko 4.9 smo pridobili iz znanih podatkov, ki smo jih obdelali z metodami opisanih v tem poglavju. Slika predstavlja končni rezultat dela. Z grafom si lahko pomagamo razporejati različne vremenske pojave. Zgoraj opisane točke A, B in C predstavljajo le eno izmed možnosti uporabe grafa.

Vztrajne točke, ki ustrezajo pogojem za nastanek toče, pripadajo slikam z veliko povezanimi komponentami in veliko vrednostjo odbojnosti. Vendar vse te točke ne ustrezajo pogojem za točo. Razlog za to je, da pridobljena parametra (trajanje in vrednost odbojnosti) nista edina, ki vplivata na končno stanje vremena. Ugotovili smo, da jih lahko uporabimo za dodatno informacijo pri klasifikaciji vremenskih pojavov. Z upoštevanjem dodatnih vremenskih podatkov, kot so veter, temperatura, relief in podobno, bi lahko izboljšali klasifikacijo.

Vztrajnost povezanih komponent in druge metode za vizualizacijo in analizo vremenskih podatkov prikazane v tem delu, so, kolikor nam je znano, nove. Z nadaljnjo izboljšavo algoritma bi lahko omogočili napovedovanje vremena. S tem bi lahko predvidevali, kakšne bodo vrednosti novih nastajajočih vlažnih gmot. Tako bi lahko vnaprej določili potencialne komponente in gmote ekstremnih situacij za nastanek toče ali poplav. Te informacije bi lahko uporabili uslužbenci ARSO in drugi delavci na tem področju za zaneslivejše napovedovanje vremena.

Poglavje 5

Sklep

V diplomskem delu smo uporabili več različnih metod za analizo in vizualizacijo vremenskih podatkov, ki temeljijo na topološki analizi oblakov na radarskih slikah. Izhodišče je bila ideja, da ni pomembna oblika podatkov, vendar nas zanimajo njihove lastnosti, kot so število komponent zračnih gmot s povečano odbojnostjo in njihova dinamika skozi čas.

Sprva smo podatke obdelali na precej enostaven način. Radarske slike smo filtrirali glede na stopnjo odbojnosti in izrisali črtne kode, ki ponazarjajo, kako se ob padajoči vrednosti praga pojavljajo in združujejo povezane komponente vlažnih gmot na eni sliki. S pomočjo zbranih informacij smo postavili hipotezo o tem, kako se iz podatkov o vztrajnosti zračnih gmot razbere, da je na nekem območju nevarnost toče.

Vnaslednjem koraku smo se osredotočili na območja z visokimi odbojnimi vrednostmi na radarskih slikah, s katerimi bi lahko zaznali ekstremne vremenske pojave. Iz vztrajnih diagramov in pripadajočih črtnih kod pri podrobnejši filtraciji smo pridobili informacije o številu in trajanju komponent in o tem, kako hitro se odbojnost spreminja pri natančnejši resoluciji.

Najbolj zanimive rezultate smo videli v času trajanja komponent in njihovih maksimalnih odbojnih vrednostih, zato smo izpostavili ta dva parametra in ju prikazali na novem dvodimenzionalnem grafu. Z uporabo hipoteze smo na grafu označili območje, ki vsebuje povezane komponente, za katere

menimo, da so potencialni kandidati za točo. Ta dva parametra se lahko uporabita kot eden izmed klasifikatorjev pri napovedovanju vremenskih pojavov.

Za natančnejše napovedovanje razmer za točo in druge ekstremne vremenske podatke bi morali uporabiti še vrsto drugih vremenskih in geografskih parametrov. Tak izboljšan model bi bilo nato dobro povezati z nekim smiselnim strojnim modelom učenja, ki bi se na podanih podatkih naučil napovedovati ali pa opozoriti na morebitne kandidate ekstremnih vremenskih pojavov.

Literatura

- [1] I. K. Darcy, *Applied Topology*, Department of Mathematics, University of Iowa. Dostopno na:
<http://homepage.math.uiowa.edu/~idarcy/AppliedTopology.html>
- [2] P. Saveliev, *Topology Illustrated*, Marshall University, pogl. 1.2, 3.2. Dostopno na:
http://inperc.com/wiki/index.php?title=Topology_Illustrated
- [3] H. Edelsbrunner, D. Morozov, *Persistent Homology: Theory and Practice*, Mathematics Subject Classification, 2010. Dostopno na:
<http://pub.ist.ac.at/~edels/Papers/2012-P-11-PHTheoryPractice.pdf>
- [4] H. Adams, A. Tausz, *JavaPlex Tutorial*, Dostopno na:
<https://code.google.com/p/javaplex/wiki/Tutorial>
- [5] H. Edelsbrunner, J. Harer, *Computational topology*, Department of Computer Science and Mathematics, Duke University, North Carolina, 2008.
- [6] G. Carlsson, A. Zomorodian, A. Collins, L. Guibas, *Persistence barcodes for shapes*, Intl. J. Shape Modeling, 2005, str. 149-187.
- [7] R. Ghrist, *Barcodes: the persistent topology of data*, Bull. Amer. Math. Soc. 45, 2008, str. 61-75.
- [8] H. Edelsbrunner, D. Letscher, A. Zomorodian, *Topological persistence and simplification*, Discrete Comput. Geom. 28, št. 4, 2002, str. 511-533.